

## 9장. 예측



# 예측

---

- 시계열 분석의 핵심
- 과거의 관측값으로 가장 설명력이 높은 모형 추정
- 추정된 모형으로 외삽법(extrapolation method)에 의한 미래 유추
- 모형식별, 모형추정 및 모형진단 단계를 거쳐 최종 선택된 모형으로 예측 진행

## 예측값과 예측오차

---

- $Z_n(l)$  : 과거의 자료인  $Z_1, \dots, Z_n$ 을 이용한  $Z_{n+l}$ 의 예측값
  - 시점  $n$ 으로부터  $l$ -시차 후의 예측
  - $n$ : 원점(origin),  $l$ : 선행시차(lead time)

- 예측오차 :

$$e_l = Z_{n+l} - Z_n(l)$$

- 예측 정확성 측도(Accuracy measure): MAE, RMSE, MAPE, MASE 등
- 예측모형의 선택: 예측 정확성 측도가 최소가 되는 모형

## 예측 정확성 측도

---

- 자료의 크기에 영향을 받는 측도
  - Mean absolute error(MAE) =  $mean(|e_l|)$
  - Root mean squared error(RMSE) =  $\sqrt{mean(e_l^2)}$
- 자료의 크기에 영향을 받지 않는 측도
  - Mean absolute percentage error(MAPE) =  $mean\left(100 \times \frac{e_l}{Z_{n+l}}\right)$
  - Mean absolute scaled error(MASE) =  $mean\left(\left|\frac{e_l}{MAE_{naive}}\right|\right)$
- 일반적으로 많이 사용되는 예측모형 선택 기준
  - MASE가 최소인 모형 선택

## 궁극적 예측함수와 예측구간

---

- 궁극적 예측함수: 큰 값의  $l$  에 대한  $z_n(l)$ 의 변화를 나타내는 함수
- 예측구간:
  - $l$  이 커짐에 따라 예측구간의 폭은 일단 점점 넓어짐
  - 일정 시차 이후
    - 정상 시계열의 경우: 그 폭이 점차 일정한 값으로 수렴
    - 비정상 시계열의 경우: 계속 넓어지는 현상 발생

## R에서의 예측

---

- 패키지 forecast의 함수 forecast( )로 시행

`forecast(object, h=, level= )`

object: 함수 `arima( )`, `Arima( )`로 생성된 객체

h : 예측하고자 하는 선행시차

level : 예측신뢰수준

## 예제 1: 정상 시계열 모형에 대한 예측

---

- AR(1), MA(1), ARMA(1,1) 모형에서
  - 각각 200개의 난수를 발생
  - 처음 180개의 데이터(training data)로 모형 적합
  - 나머지 20개 데이터(test data)에 대한 예측 실시 후 예측 정확성 측도 계산
  - 원자료, 예측값 및 예측구간을 하나의 그래프에 표현

## 1) AR(1)의 경우

---

- 자료의 생성 및 모형 적합:  $Z_t = 0.7Z_{t-1} + \varepsilon_t$

```
> set.seed(1234) # 같은 자료 확보를 위한 것  
> z <- arima.sim(n=200,list(ar=.7))  
> z1 <- window(z,end=180); t1 <- window(z,start=181)  
> fit_z1 <- arima(z1,order=c(1,0,0),include.mean=FALSE)
```

- 예측 실시 및 결과 확인

```
> fore_z1 <- forecast(fit_z1,h=20)
```

```
> summary(fore_z1)
```



- 예측 정확성 측도

```
> accuracy(fore_z1,t1)
```

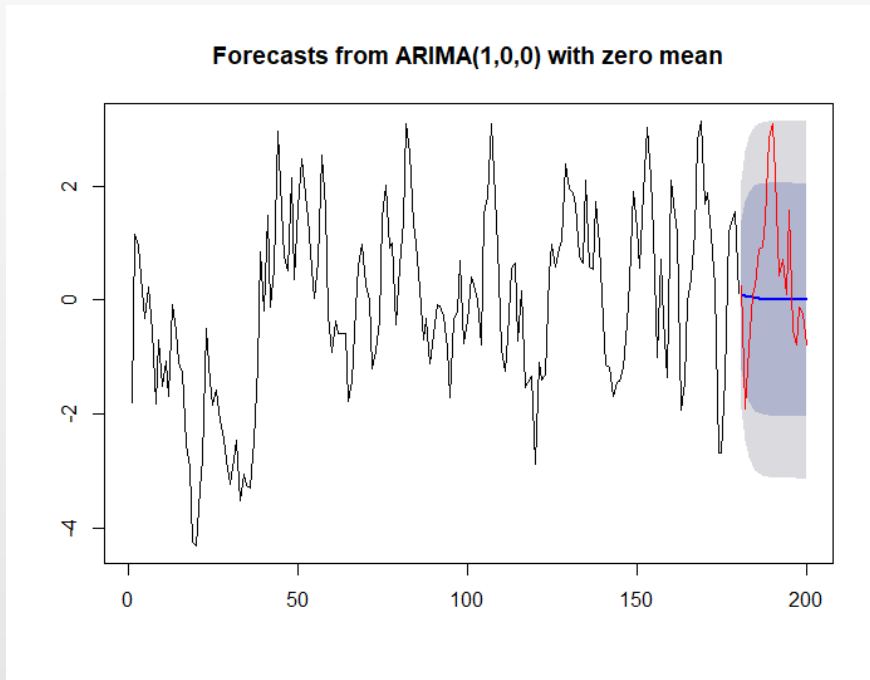
	ME	RMSE	MAE	MPE	MAPE
Training set	-0.01349832	1.021939	0.8026364	-82.38506	251.98441
Test set	0.45906685	1.327150	0.9891131	95.34048	95.34048

	MASE	ACF1	Theil's U
Training set	0.928592	0.06152305	NA
Test set	1.144332	0.63497786	1.064339

- 그래프 작성

```
> plot(fore_z1)  
> lines(181:200,t1,col="red")
```



- 예측값은 자료의 평균 0으로 수렴
- 예측구간은 점차 넓어지다 일정한 간격 유지
- 디폴트 예측구간 신뢰수준: 80%, 95%

## 2) MA(1)의 경우

- 자료 생성 및 모형 적합:  $Z_t = \varepsilon_t + 0.7\varepsilon_{t-1}$

```
> set.seed(1234)
> z <- arima.sim(n=200,list(ma=.7))
> z2 <- window(z,end=180); t2 <- window(z,start=181)
> fit_z2 <- arima(z2,order=c(0,0,1),include.mean=FALSE)
> fore_z2 <- forecast(fit_z2,h=20)
```

- 예측 정확성 측도

```
> accuracy(fore_z2,t2)
```

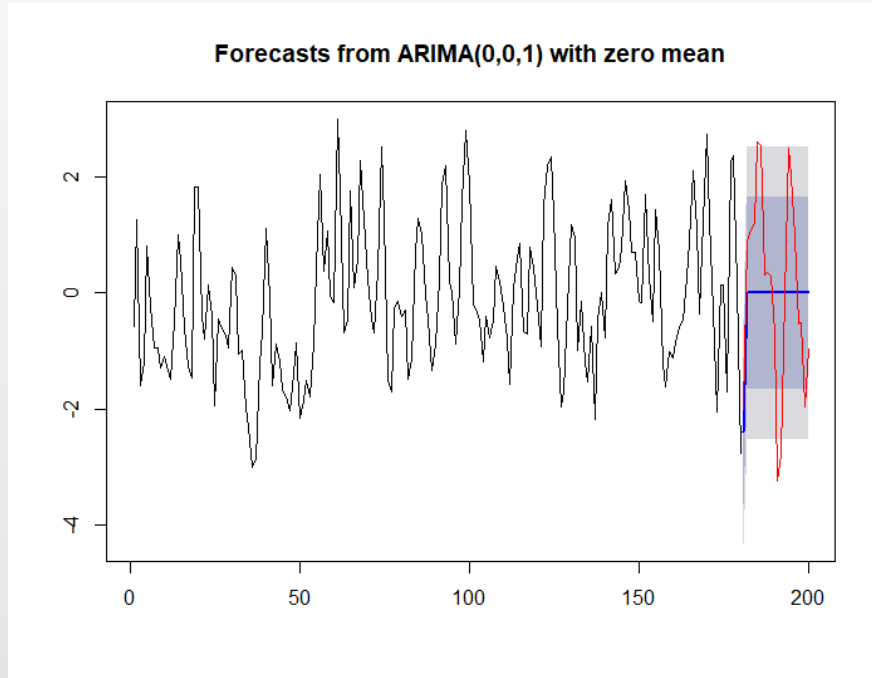
	ME	RMSE	MAE	MPE	MAPE
Training set	-0.0676195	0.987859	0.7770894	104.97235	211.00077
Test set	0.2236632	1.622996	1.2978771	94.05758	95.94242

	MASE	ACF1	Theil's U
Training set	0.8054596	-0.03175746	NA
Test set	1.3452604	0.63951057	0.9801794

- 그래프 작성

```
> plot(fore_z2)  
> lines(181:200,t2,col="red")
```



- 1시차 후 예측만 유효한 결과이며, 그 이후에는 그냥  $\mu$ 가 됨
- MA(q) 모형은 q-시차 후 예측까지만 가능

### 3) ARMA(1,1)의 경우

---

- 모형:  $Z_t = 0.7Z_{t-1} + \varepsilon_t + 0.5\varepsilon_{t-1}$

```
> set.seed(1234)
> z <- arima.sim(n=200,list(ar=.7,ma=.5))
> z3 <- window(z,end=180); t3 <- window(z,start=181)
> fit_z3 <- arima(z3,order=c(1,0,1),include.mean=FALSE)
> fore_z3 <- forecast(fit_z3,h=20)
> accuracy(fore_z3,t3)
```

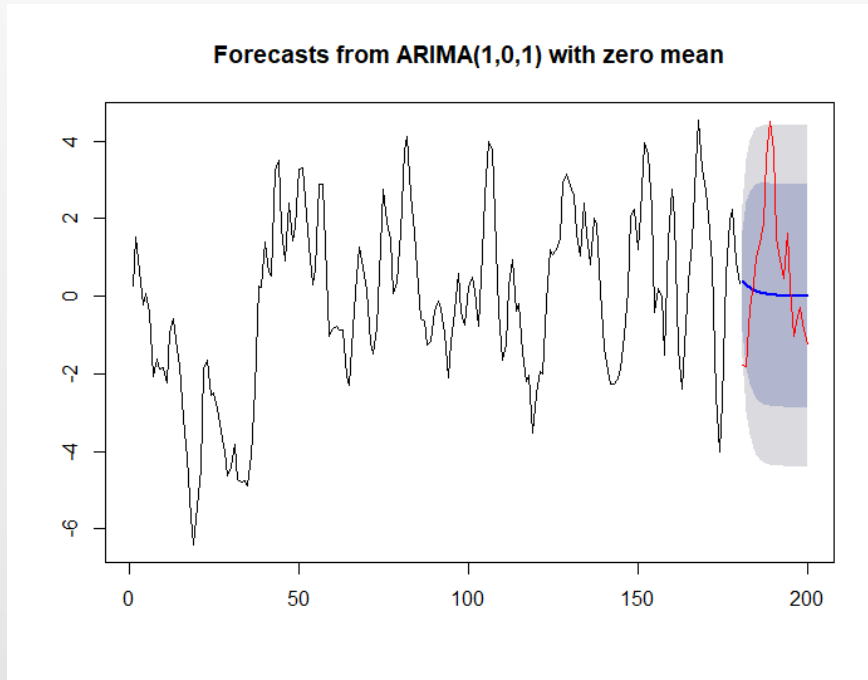
	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	-0.0153926	1.00333	0.787043	-19.6315	138.602	0.8355
Test set	0.5871135	1.87877	1.463298	101.0585	101.058	1.5535

	ACF1	Theil's U
Training set	0.0206307	NA
Test set	0.7631018	1.119702

- 그래프 작성

```
> plot(fore_z3)  
> lines(181:200,t3,col="red")
```



AR 모형과 비슷한 양상

## 예제 2: 비정상 시계열 모형에 대한 예측

---

- ARIMA(1,1,1) 모형
  - 절편이 있는 모형과 절편이 없는 모형을 구분하여
  - 200개의 난수를 발생
  - 처음 180개의 데이터(training data)로 모형 적합
  - 나머지 20개 데이터(test data)에 대한 예측 실시 후 예측 정확성 측도 계산
  - 원자료, 예측값 및 예측구간을 하나의 그래프에 표현

## 1) 절편이 없는 ARIMA(1,1,1) 모형

---

- 자료 생성 및 모형 적합:  $(1 - 0.7B)(1 - B)Z_t = (1 + 0.5B)\varepsilon_t$

```
> set.seed(1234)
> z <- arima.sim(n=200,list(order=c(1,1,1),ar=.7,ma=.5))
> z <- as.ts(z[-1])
> z4 <- window(z,end=180); t4 <- window(z,start=181)
> fit_z4 <- Arima(z4,order=c(1,1,1))
```

- `arima.sim()`:  $d=1$ 의 경우, 0이 자료의 첫 번째 값으로 추가되어  $n+1$ 개 자료 생성
- `arima.sim()`으로 생성된  $z$ 는 `ts` 객체
- `z[-1]`: 숫자형 벡터



- 예측 및 그래프 작성

```
> fore_z4 <- forecast(fit_z4,h=20)
> accuracy(fore_z4,t4)
```

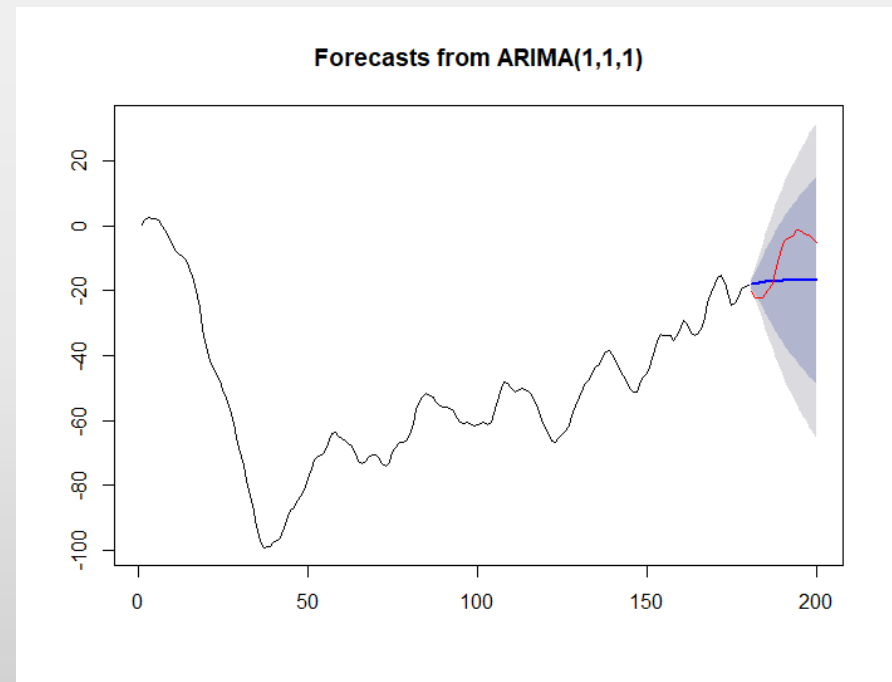
	ME	RMSE	MAE	MPE	MAPE
Training set	-0.0169358	0.999159	0.783228	3.44312	6.52263
Test set	6.8109400	10.462696	9.095907	-298.4846	309.23358

	MASE	ACF1	Theil's U
Training set	0.4255067	0.02071345	NA
Test set	4.9415586	0.93241724	15.51268

```
> plot(fore_z4)
> lines(181:200,t4,col="red")
```

- 예측값은 특정한 값으로 수렴
- 예측구간은 계속 넓어짐

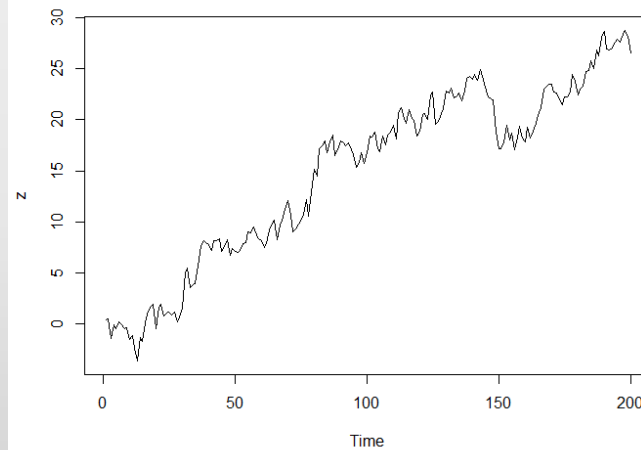
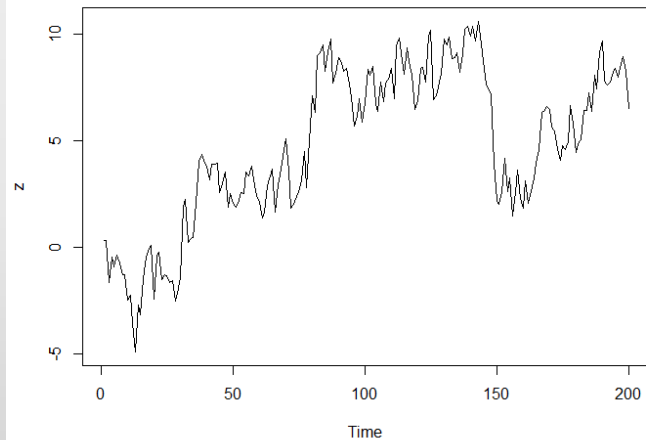


## 2) 절편이 있는 ARIMA(1,1,1)의 경우

- 절편이 있는 ARIMA(1,1,1) 자료 생성:

```
> set.seed(1235)
> z <- arima.sim(n=200, list(order=c(1,1,1), ar=.7, ma=-.7))
> z <- as.ts(z[-1])
```

```
> nobs <- length(z)
> z <- z+0.1*(1:nobs)
```



- 모형 적합 및 예측

```
> z5 <- window(z,end=180); t5 <- window(z,start=181)
> fit_z5 <- Arima(z5,order=c(1,1,1),include.drift=TRUE)
```

```
> fore_z5 <- forecast(fit_z5,h=20)
> accuracy(fore_z5,t5)
```

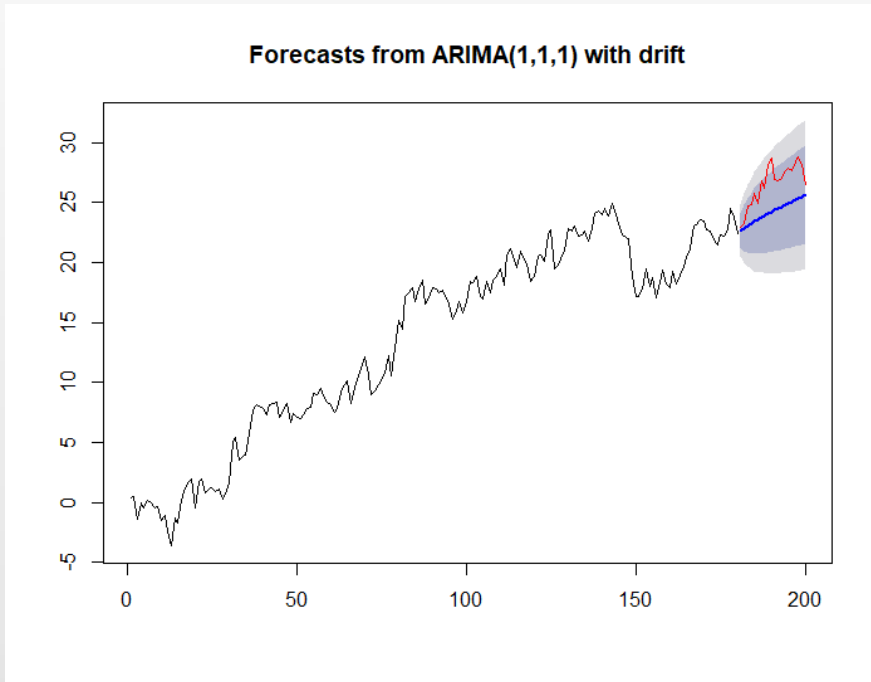
	ME	RMSE	MAE	MPE	MAPE
Training set	-0.00704728	1.08012	0.829808	10.40168	41.09658
Test set	2.35968033	2.58670	2.359680	8.67231	8.67231

	MASE	ACF1	Theil's U
Training set	0.9567794	-0.02849251	NA
Test set	2.7207399	0.45765804	2.585701

## - 그래프 작성

```
> plot(fore_z5)  
> lines(181:200,t5,col="red")
```



- 예측값 지속적 상승: 절편의 존재로 인한 추세 반영
- 예측구간은 계속 넓어짐

### 예제 3: 연습문제 8.2

---

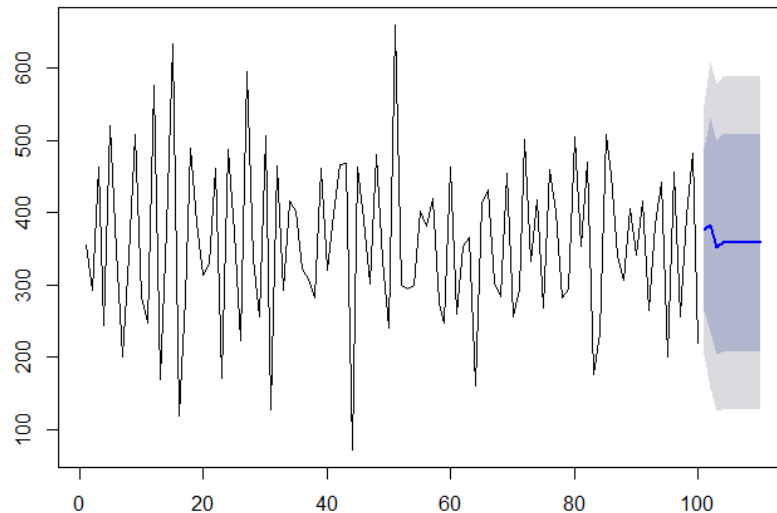
- 다음의 자료에 대하여 8장에서 최종 선택한 모형을 이용하여 10 선행 시차까지의 예측을 실시하라.
- 데이터:
  - ex8\_2a.txt
  - ex8\_2b.txt

## ex8\_2a.txt

최종 모형: ma2항이 제거된 절편이 있는 MA(3)

```
> ex8_2a <- scan("D:/Data/ex8_2a.txt")  
> fit_a <- Arima(ex8_2a, order=c(0,0,3), fixed=c(NA,0,NA,NA))  
> summary(forecast(fit_a))  
> plot(forecast(fit_a))
```

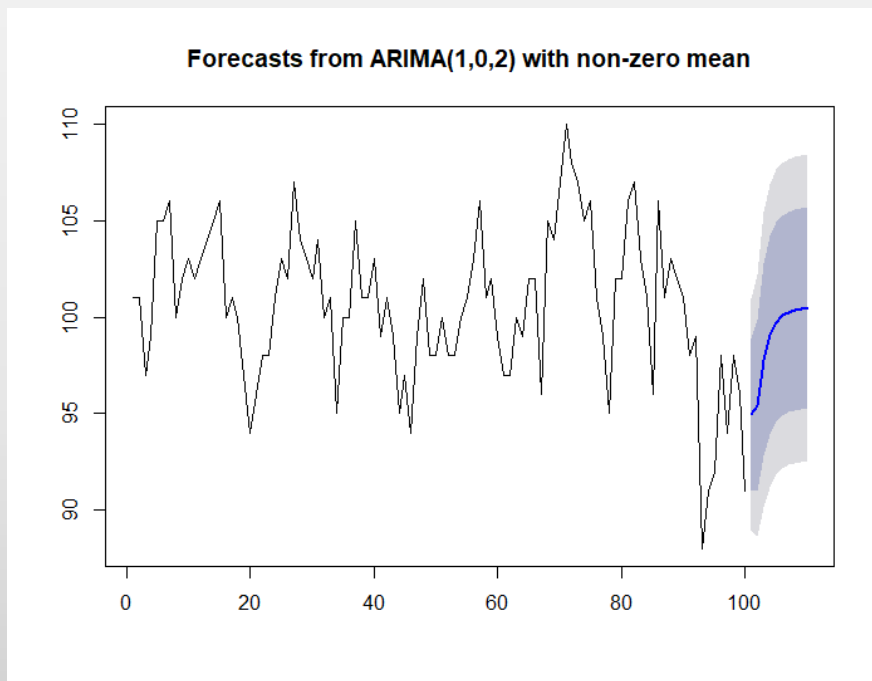
Forecasts from ARIMA(0,0,3) with non-zero mean



## ex8\_2b.txt

최종 모형: ma1항이 제거된 절편이 있는 ARMA(1,2)

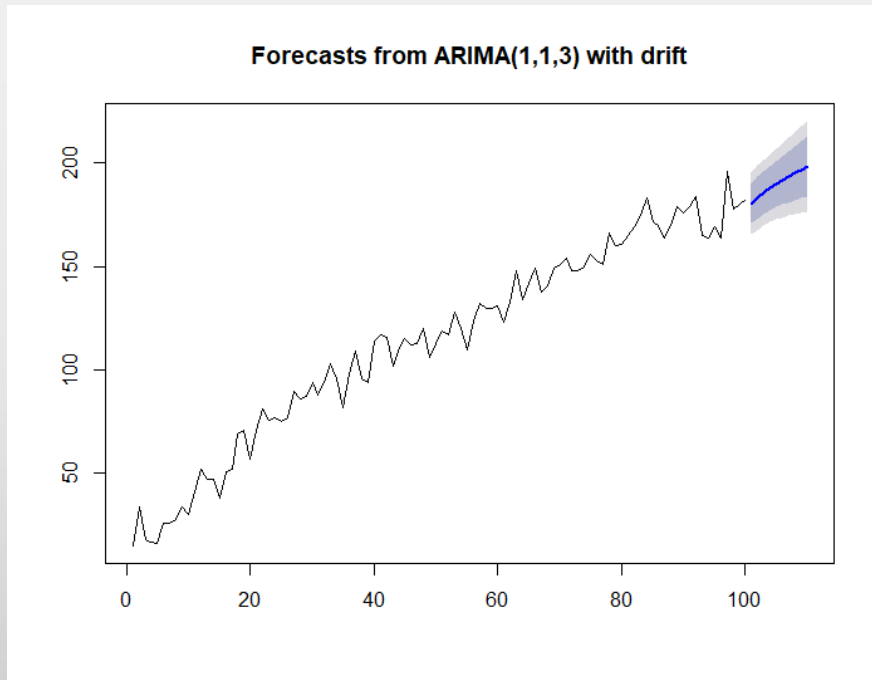
```
> ex8_2b <- scan("D:/Data/ex8_2b.txt")  
> fit_b <- arima(ex8_2b, order=c(1,0,2), fixed=c(NA,0,NA,NA))  
> summary(forecast(fit_b))  
> plot(forecast(fit_b))
```



## 예제 4: 연습문제 7.5 (d)

최종 모형: 절편이 유의하고 ma2 항이 제거된 ARIMA(1,1,3)

```
> ex7_5d <- scan("D:/Data/ex7_5d.txt")
> fit_d <- Arima(ex7_5d, order=c(1,1,3), include.drift=TRUE,
                 fixed=c(NA, NA, 0, NA, NA))
> summary(forecast(fit_d))
> plot(forecast(fit_d))
```





## 예제 5: interest.txt

최종 모형: 절편이 없는  $ARI(1,1)$

```
> rate <- scan("D:/Data/interest.txt")  
> fit_r <- Arima(rate, order=c(1,1,0))  
> summary(forecast(fit_r))  
> plot(forecast(fit_r))
```

